CLAIMS

What is claimed is:

1. 1. A method of providing information about a product, the product available for purchase from a plurality of sources, the method comprising:

   receiving a selection of a product category from a predefined set of product categories using information about the product;

   accessing a list of extraction parameters for the product category;

   receiving a selection of at least one extraction parameter in the list of extraction parameters;

   for each of the plurality of sources, creating a corresponding program including

   identifying a corresponding web site, the corresponding web site selling the product and

   providing a tool for creating the corresponding program to extract data from the web site using the at least one extraction parameter;

   receiving a connection from a client, the connection including a value for the at least one extraction parameter; and

   for each of the plurality of sources in the product category, providing product information for the product using the value for the at least one extraction parameter and the corresponding program.

2. 2. The method of claim 1, wherein the providing the tool for creating the corresponding program to extract data from the corresponding web site using the at least one extraction parameter further comprises providing a graphical user interface tool for developing the corresponding program, the graphical user interface tool including a web browser.
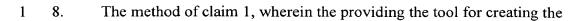
1   3.   The method of claim 2, wherein the graphical user interface tool further includes a

2   first tool, the first tool for developing an extraction pattern, the extraction pattern

3   identifying a plurality of portions of a document on the corresponding web site.

1   4.   The method of claim 3, wherein the graphical user interface tool further comprises:

2   receiving a selection signal;

3   applying the extraction pattern to find a matching pattern in a document displayed

4   in a source view in the web browser;

5   displaying a rendered version of the matching pattern in a window.

1   5.   The method of claim 3, wherein the graphical user interface tool further includes a

2   plurality of predefined extraction patterns.

1   6.   The method of claim 5, wherein the plurality of predefined extraction patterns

2   include at least one of an extraction pattern for matching a hyperlink, an extraction pattern

3   for matching a form, and an extraction pattern for matching a price.

1   7.   The method of claim 3, wherein the graphical user interface tool further comprises:

2   identifying a form on the document on the corresponding web site;

3   creating a step in the corresponding program, the step to submit the form without

4   retrieving the document;

5   generating a plurality of parameters associated with the step, the plurality of

6   parameters corresponding to inputs in the form; and

7   identifying at least one of the plurality of parameters with the at least one

8   extraction parameter.

1   8.   The method of claim 1, wherein the providing the tool for creating the

2   corresponding program to extract data from the corresponding web site using the at least

3   one extraction parameter further comprises defining a plurality of steps wherein at least

4   one step in the plurality of steps interacts with the corresponding web site and operates on

5   the results of the interaction.

1   9.   The method of claim 8, wherein the defining the plurality of steps comprises for

2   each of the plurality of steps, receiving a selection of an extraction command from a

3   predetermined list of extraction commands.

1   10.   The method of claim 9, wherein the predetermined list of extraction commands

2   includes extraction commands for retrieving multiple matches of an extraction pattern

3   from a document.

1   11.   The method of claim 9, wherein the predetermined list of extraction commands

2   includes extraction commands for extracting data from a first document and a second

3   document, the first document including a reference to the second document.

1   12.   The method of claim 9, wherein at least one step in the plurality of steps includes a

2   test condition comprising a logical test for at least one corresponding argument and a first

3   step in the plurality of steps, and wherein the program continues executing at the first step

4   if the logical test is satisfied.

1   13.   The method of claim 12, wherein the at least one corresponding argument includes

2   an extraction pattern.

14.   The method of claim 12, wherein the test condition further comprises a result code,

wherein the program returns an error if the result code is a web site changed result code.

1 15. The method of claim 12, wherein the test condition further comprises a result code,

2 wherein the program returns an error if the result code is a no matching products result

3 ~~code.~~
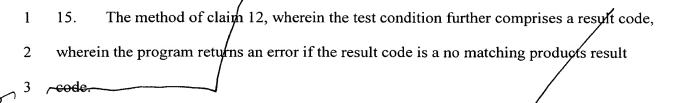
1 16. The method of claim 9, wherein the predetermined list of extraction commands

2 includes extraction commands for segmenting a document into a plurality of units, each of

3 the plurality of units matching an extraction pattern.

1 17. The method of claim 16, wherein at least one step in the plurality of steps uses an

2 extraction command to segment a document into a plurality of units, and wherein the step

3 further includes a test condition, the test condition comprising a logical test and at least

4 one argument, and wherein for each of the plurality of units, the logical test is computed

5 with the at least one argument, and the unit is removed from the plurality of units if the

6 logical test is not satisfied with the at least one argument.

1 18. An apparatus for providing information about a product, the product available for

2 purchase from a plurality of sources, the apparatus comprising:

3 means for receiving a selection of a product category from a predefined set of

4 product categories using information about the product;

5 means for accessing a list of extraction parameters for the product category;

6 means for receiving a selection of at least one extraction parameter in the list of

7 extraction parameters;

8 means for creating a corresponding program for each of the plurality of sources,

9 the means for creating a corresponding program for each of the plurality of

10 sources including·

12   means for identifying a corresponding web site, the corresponding web site

   selling the product and

13   means for creating the corresponding program to extract data from the web site

14   using the at least one extraction parameter;

15   means for receiving a connection from a client, the connection including a value

16   for the at least one extraction parameter; and

17   means for providing product information for the product from each of the plurality

18   of sources using the value for the at least one extraction parameter and the

19   corresponding program.

1   19.   The apparatus of claim 18, wherein the means for creating a corresponding

2   program to extract data from the web site includes means for selecting an instruction from

3   a predetermined list of instructions.

1   20.   The apparatus of claim 18, wherein the means for creating a corresponding

2   program to extract data from the web site includes means for developing an extractor

3   pattern in a web browser.

1   21.   A computer data signal embodied in a carrier wave comprising:

2   a computer program for developing descriptions of data of interest

3   a set of instructions for developing an extractor pattern interactively in a web

4   page;

5   a set of instructions for receiving a selection of an instruction from a predefined

6   set of instructions for inclusion of the instruction in the description of data

7   of interest;

a set of instructions for associating the extractor pattern with the instruction; and

a set of instructions for testing the instruction using the extractor pattern and the contents of a buffer;

22.   An apparatus comprising a computer, the computer comprising a processor and a memory, the memory including a plurality of descriptions of data of interest, the processor running a program the program accepting an input and generating an output, the input identifying a subset of the plurality of descriptions of data of interest and a plurality of values for a plurality of extraction parameters, the output including data of interest retrieved from a plurality of web sites corresponding to data of interest matching the plurality of values for the plurality of extraction parameters at each of the plurality of web sites corresponding to the subset of plurality of descriptions of data of interest.